Amendments to the Specification:

Please replace the paragraph on page 2, line 6 with the following amended paragraph:

This application is related to U.S. Application Serial No. 09/735,743, filed December 12, 2000, ~~Attorney Docket No. 3298.1,~~ which is incorporated herein by reference in its entirety for all purposes.

Please replace the paragraph on page 3, line 19 with the following amended paragraph:

Normalization is often, but not always, a necessary and fundamental step for comparison of results from two or more probe arrays. A normalization factor (f) is used to adjust signals from probe arrays (*e.g.*, intensity values) to compensate for array to array variations or variations due to other factors, such as sample preparation. If $I^{(1)}$ [[be]] is the intensity from a first probe array, $I^{(2)}$ from a second probe array and the normalization factor f is such that $I^{(1)}$ and $fI^{(2)}$ are comparable.

Please replace the paragraph on page 5, line 1 with the following amended paragraph:

Computer implemented methods for comparing the expression of a gene in a first sample with a second sample are also provided. The methods may include steps of providing a first plurality of intensity values ($I_i^{(1)}$), each of which reflects the expression of the gene in the first sample, where the intensity values are obtained from a first nucleic acid probe array; providing a second plurality of intensity values ($I_i^{(2)}$), each of which reflects the expression of the gene in the second sample, wherein the intensity values are obtained from a second nucleic acid probe array; calculating a p-value using one-sided Wilcoxon's signed rank test, wherein the p-value is for a null hypothesis that

median(f(x) $I_i^{(2)} - I_i^{(1)}$)=0 and an alternative hypothesis that ~~*median((f(x)* $I_i^{(1)} - I_i^{(2)}$)>0,~~ *median(f(x)* $I_i^{(1)} - I_i^{(2)}$)>0, wherein said *f(x)* is a normalization factor; and indicating whether the expression of gene is increased in the second sample in comparison with the first sample based upon said *p*-value. ~~The each~~ Each of the intensity values may be from one probe (such as a probe that is designed to target the transcript of the gene) on the nucleic acid probe arrays.

Please replace the paragraph on page 6, line 6 with the following amended paragraph:

In another aspect of the invention, computer software products and systems for performing the methods of the invention are also provided. The computer software product ~~include~~ includes code for performing the steps of the method of the invention and a computer readable medium for storing the code. A system of the invention ~~include~~ includes a processor; and a memory being coupled with the processor, the memory storing a plurality of machine instructions that cause the processor to perform the method steps of the invention.

Please replace the paragraph on page 6, line 15 with the following amended paragraph:

The accompanying drawings, which are incorporated in and form a part of this specification, illustrate embodiments of the invention and, together with the description, serve to explain the principles of the invention:

Figure 1 illustrates an example of a computer system that may be utilized to execute the software of an embodiment of the invention.

Figure 2 illustrates a system block diagram of the computer system of Fig. 1.

Figure 3 shows a computerized process for comparative analysis of gene expression data from two probe arrays.

Figure 4 shows 2 fold detected rate and no change error rate for 9912072, 9913514 and 9914059.

Figure 5 shows 2 fold detected rate and no change error rate for 9912072BG, 9913514BG and 9914059BG.

Please replace the paragraph on page 8, line 14 with the following amended paragraph:

"A target molecule" refers to a biological molecule of interest. The biological molecule of interest can be a ligand, receptor, peptide, nucleic acid (oligonucleotide or polynucleotide of RNA or DNA), or any other of the biological molecules listed in U.S. Patent No. 5,445,934 at col. 5, line 66 to col. 7, line 51. For example, if transcripts of genes are the interest of an experiment, the target molecules would be the transcripts. Other examples include protein fragments, small molecules, etc. "Target nucleic acid" refers to a nucleic acid (often derived from a biological sample) of interest. Frequently, a target molecule is detected using one or more probes. As used herein, a "probe" is a molecule for detecting a target molecule. It can be any of the molecules in the same classes as the target referenced above. A probe may refer to a nucleic acid, such as an oligonucleotide, capable of binding to a target nucleic acid of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation. As used herein, a probe may include natural (i.e. A, G, U, C, or T) or modified bases (7-deazaguanosine, inosine, etc.). In addition, the bases in probes may be joined by a linkage other than a phosphodiester

bond, so long as the bond does not interfere with hybridization. Thus, probes may be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages. Other examples of probes include antibodies used to detect peptides or other molecules, any ligands for detecting its binding partners. When referring to targets or probes as nucleic acids, it should be understood that these are illustrative embodiments that are not to limit the invention in any way.

Please replace the paragraph on page 11, line 1 with the following amended paragraph:

Typically, a nucleic acid sample is a labeled with a signal moiety, such as a fluorescent label. The sample is hybridized with the array under appropriate conditions. The arrays are washed or otherwise processed to remove non-hybridized sample nucleic acids. The hybridization is then evaluated by detecting the distribution of the label on the chip. The distribution of label may be detected by scanning the arrays to determine florescence intensities distribution. Typically, the hybridization of each probe is reflected by several pixel intensities. The raw intensity data may be stored in a gray scale pixel intensity file. The GATC™ Consortium has specified several file formats for storing array intensity data. The final software specification is available at ~~www.gatcconsortium.org~~ the website of the GATC™ Consortium and is incorporated herein by reference in its entirety. The pixel intensity files are usually large. For example, a GATC™ compatible image file may be approximately 50 Mb if there are about 5000 pixels on each of the horizontal and vertical axes and if a two byte integer is used for every pixel intensity. The pixels may be grouped into cells (see, GATC™ software specification). The probes in a cell are designed to have the same sequence (i.e.,

each cell is a probe area). A CEL file contains the statistics of a cell, e.g., the 75 percentile and standard deviation of intensities of pixels in a cell. The 75 percentile of pixel intensity of a cell is often used as the intensity of the cell. Methods for signal detection and processing of intensity data are additionally disclosed in, for example, U.S. Patents Numbers 5,547,839, 5,578,832, 5,631,734, 5,800,992, 5,856,092, 5,936,324, 5,981,956, 6,025,601, 6,090,555, 6,141,096, 6,141,096, and 5,902,723. Methods for array based assays, computer software for data analysis and applications are additionally disclosed in, e.g., U.S. Patent Numbers 5,527,670, 5,527,676, 5,545,531, 5,622,829, 5,631,128, 5,639,423, 5,646,039, 5,650,268, 5,654,155, 5,674,742, 5,710,000, 5,733,729, 5,795,716, 5,814,450, 5,821,328, 5,824,477, 5,834,252, 5,834,758, 5,837,832, 5,843,655, 5,856,086, 5,856,104, 5,856,174, 5,858,659, 5,861,242, 5,869,244, 5,871,928, 5,874,219, 5,902,723, 5,925,525, 5,928,905, 5,935,793, 5,945,334, 5,959,098, 5,968,730, 5,968,740, 5,974,164, 5,981,174, 5,981,185, 5,985,651, 6,013,440, 6,013,449, 6,020,135, 6,027,880, 6,027,894, 6,033,850, 6,033,860, 6,037,124, 6,040,138, 6,040,193, 6,043,080, 6,045,996, 6,050,719, 6,066,454, 6,083,697, 6,114,116, 6,114,122, 6,121,048, 6,124,102, 6,130,046, 6,132,580, 6,132,996 and 6,136,269, all of which are incorporated by reference in their entireties for all purposes.

Please replace the paragraph on page 14, line 1 with the following amended paragraph:

The embodiments of the invention will be described using GeneChip® high oligonucleotide density probe arrays (available from Affymetrix, Inc., Santa Clara, CA, USA) as exemplary embodiments. One of skill in the art would appreciate that the embodiments of the invention are not limited to high density oligonucleotide probe

arrays. In contrast, the embodiments of the invention are useful for analyzing any parallel large scale biological analysis, such as those using nucleic acid probe array, protein arrays, etc.

Please replace the paragraph on page 14, line 13 with the following amended paragraph:

In the preferred embodiment, oligonucleotide probes are synthesized directly on the surface of the array using photolithography and combinatorial chemistry as disclosed in several patents previous incorporated by reference. In such embodiments, a single square-shaped feature on an array contains one type of probe. Probes are selected to be specific against desired target. Methods for selecting probe sequences are disclosed in, for example, U.S. Patent Application Nos. 09/718,295, ~~Attorney Docket Number 3359;~~ filed November 21, 2000, ~~Attorney Docket Number 3367~~ 09/721,042, filed November 21, 2000, and 60/252,617, ~~Attorney Docket Number 3373,~~ filed November 21, 2000, all incorporated herein by reference in their entireties for all purposes.

Please replace the paragraph on page 17, line 15 with the following amended paragraph:

Figure 1 illustrates an example of a computer system that may be used to execute the software of an embodiment of the invention. Figure 1 shows a computer system 1 that includes a display 3, screen 5, cabinet 7, keyboard 9, and mouse 11. Mouse 11 may have one or more buttons 13 for interacting with a graphic user interface. Cabinet 7 houses a CD-ROM or DVD-ROM drive [[13]] 15, system memory and a hard drive (*see* Figure 2) which may be utilized to store and retrieve software programs incorporating computer code that implements the invention, data for use with the invention and the like.

Although a CD [[15]] 17 is shown as an exemplary computer readable medium, other computer readable storage media including floppy disk, tape, flash memory, system memory, and hard drive may be utilized. Additionally, a data signal embodied in a carrier wave (*e.g.*, in a network including the Internet) may be the computer readable storage medium.

Please replace the paragraph on page 18, line 4 with the following amended paragraph:

Figure 2 shows a system block diagram of computer system 1 used to execute the software of an embodiment of the invention. As in Figure 1, computer system 1 includes monitor 3, keyboard 9, and mouse 11. Computer system 1 further includes subsystems such as a central processor [[51]] 50, system memory [[53]] 52, fixed storage [[55]] 60 (*e.g.*, hard drive), removable storage [[57]] 58 (*e.g.*, CD-ROM), display adapter [[59]] 56, sound card [[61]], speakers [[63]] 64, and network interface [[65]] 62. Other computer systems suitable for use with the invention may include additional or fewer subsystems. For example, another computer system may include more than one processor [[51]] 50 or a cache memory. Computer systems suitable for use with the invention may also be embedded in a measurement instrument.

Please replace the paragraph on page 21, line 1 with the following amended paragraph:

Expression level controls are probes that hybridize specifically with constitutively expressed genes in the biological sample. Virtually any constitutively expressed gene provides a suitable target for expression level controls. Typically expression level control probes have sequences complementary to subsequences of constitutively

expressed "housekeeping genes" including, but not limited to the ß-actin gene, the transferrin receptor gene, the GAPDH gene, and the like. Housekeeping genes, or maintenance genes, are those genes constitutively expressed to maintain cellular function (See, Watson, J.D., N.H. Hopkins, J.W. Roberts, J.A. Steitz, A.M. Weiner, A.M. *Molecular Biology of the Gene*, Vol.1, 1965, which is incorporated herein in its entirety by reference for all purposes). ~~U.S. Patent Application Serial Number____, Attorney Docket Number 3340.1, which is incorporated herein by reference for all purposes.~~

Please replace the paragraph on page 23, line 4 with the following amended paragraph:

In one aspect of the invention, computer implemented methods for calculating a normalization factor are provided. The ~~method~~ methods include providing a first intensity value ($I^{(1)}$) of a probe in a first probe array and a second intensity value ($I^{(2)}$) of the probe in a second probe array; obtaining the geometric mean ($x = \sqrt{I^{(1)} I^{(2)}}$) of $I^{(1)}$ and $I^{(2)}$; calculating said normalization factor according to: $f(x) = e^{h(x)}$, where $h(x)$ is derived from referential intensities from the first and second probe arrays. $h(x)$ may be derived by relating geometric means (x_l') of first referential intensities ($RI_i^{(1)}$) in the first probe array and second referential intensities ($RI_i^{(2)}$) in the second probe array to:

$$y_l = \log\left(\frac{RI_i^{(1)}}{RI_i^{(2)}}\right).$$

Please replace the paragraph on page 24, line 2 with the following amended paragraph:

The normalization factor of the invention may be used to adjust for probe array to probe array variations so that intensity values from different probe arrays may be

appropriately compared. In one aspect of the invention, computer implemented methods for comparing the expression of a gene in a first sample with a second sample are also provided. One of skill in the art would appreciate that the normalization factor, methods, system and software for calculating the normalization factor of the invention are not limited to any particular method for comparison. Rather the normalization factor of the invention may be used in conjunction [[of]] with other suitable statistical comparison methods not discussed in this specification.

Please replace the paragraph on page 24, line 11 with the following amended paragraph:

In preferred embodiments, methods are provided to compare results from different probe arrays using the normalization factor of the invention. The methods may include the steps of providing a first plurality of intensity values ($I_i^{(1)}$), each of which reflects the expression of the gene in the first sample, where the intensity values are obtained from a first nucleic acid probe array; providing a second plurality of intensity values ($I_i^{(2)}$), each of which reflects the expression of the gene in the second sample, wherein the intensity values are obtained from a second nucleic acid probe array; calculating a *p*-value using one-sided Wilcoxon's signed rank test, wherein the *p*-value is for a null hypothesis that $median(f(x) I_i^{(2)} - I_i^{(1)})=0$ and an alternative hypothesis that ~~$median((f(x) I_i^{(1)} - I_i^{(2)})>0$,~~ $median(f(x) I_i^{(1)} - I_i^{(2)})>0$, wherein said $f(x)$ is a normalization factor; and indicating whether the expression of gene is increased in the second sample in comparison with the first sample based upon said *p*-value. ~~The each~~ Each of the intensity values may be from one probe (such as a probe that is designed to target the transcript of the gene) on the nucleic acid probe arrays. One of skill in the art would appreciate that

the nonparametric comparison methods of the invention are not limited to any particular normalization factors. In some embodiments, there may not be a need for any normalization factor (i.e., normalization factor=1). However, in particularly preferred embodiments, the nonparametric methods for comparative analysis employ the normalization factor of the invention.

Please replace the paragraph on page 26, line 15 with the following amended paragraph:

The performance of the method is summarized in Figures [[5]] 4 and [[6]] 5. The upper curve shows the two-fold correctly detected comparative calls using the normalization factor and non-parametric test of the invention (0 versus 0.25 pM is also included as group 0, 0.25 pM versus 0.5 pM is considered as group 1, and 0.5 pM versus 1 pM is considered as group 2, ..., 512 pM versus 1024 pM is considered as group 12). The lower curve shows the error rate of no change calls (0 pM versus 0 pM is considered as group 0, 0.25 pM versus 0.25 pM is considered as group 2,..., 1024 pM versus 1024 pM is considered as group 13.